

## Hedoniset hintamallit ja alueelliset asuntomarkkinat

Harjoitustyö 2  
Eetu Jahkonen ja Lauri Vuorio  
Maankäyttötieteiden laitos  
MAA-20.2333 Asuntomarkkinat ja arviointi  
25.5.2016

# Sisällysluettelo

1 Johdanto .....	1
2 Alueellisten hintamallien teoria ja aikaisempi tutkimus .....	2
3 Alueelliset hintamallit Helsingin aineistolla .....	4
3.1 Aineisto .....	4
3.1.1 Aineiston muodostus ja rajaukset .....	4
3.1.2 Aineiston muuttujamuunnokset ja tunnusluvut .....	4
3.2 Mallin estimointi .....	6
4 Johtopäätökset .....	9
5 Liitteet .....	10
Liite 1: Helsingin aluejako .....	10
Liite 2: Aineiston muuttujat ja muuttujamuunnokset .....	11
Liite 3: Aineiston tunnusluvut .....	12
Liite 4: Mallien residuaalien diagnostiikka .....	14
Lähdeluettelo .....	21

## Tiivistelmä

Useissa tutkimuksissa on todettu asuntomarkkinoiden olevan heterogeenisiä suurten kaupunkialueiden sisällä. Näin ollen asuntojen hintaan vaikuttavat tekijät saattavat olla erilaisia kaupunkien sisällä sijaitsevilla osamarkkinoilla. Tässä harjoitustyössä tutkittiin kerrostaloasuntojen hintaan vaikuttavia tekijöitä Helsingin kaupungin alueella estimoiden usean selittävän muuttujan regressiomalli pienimmän neliösumman menetelmällä. Tilastollisesti hyvin merkitsevien tulosten perusteella voidaan esittää Helsingin neljän eri osamarkkinan olevan hinnanmuodostukseltaan varsin erilaiset, sillä hintamallin selittävät muuttujat poikkesivat eri alueilla toisistaan. Arvokkaimman, keskusta-alueen hintamalli oli selitysasteeltaan selvästi tarkin (yli 90 %), mikä indikoi kyseisen alueen olevan selvästi homogeenisin neljästä alueesta. Toiseksi arvokkain alue vaikutti niin ikään hyvin homogeeniseltä. Tutkimuksessa ei otettu kantaa estimoitujen hintamallien tarkkuuteen yksittäisten asuntokauppojen hintojen selittäjänä. Jatkossa useita erilaisia Helsingin kaupunkialueen aluejakoja testaten olisi mahdollista tarkentaa eri osamarkkinoiden maantieteellistä rajautumista ja hintamalleja entisestään.

Avainsanat: hedoninen hintamalli, asuntomarkkinat, osamarkkina

# 1 Johdanto

Asuntojen hintateorian ja empiiristen tutkimusten perusteella voidaan osoittaa, että suurella kaupunkialueella ei ole olemassa yksiä asuntomarkkinoita, vaan hintataso on selitettävissä samoilla tekijöillä vain pienemmillä, homogeenisemmilla alueilla. Tällöin järkevämpi lähestymistapa on estimoida kaupunkialueen eri osille omat hedoniset hintamallit, joilla hinnanmuodostusta pystytään tarkastelemaan.

Hedonisten hintojen teorian näkökulmasta malleihin tulisi sisällyttää kaikki kuluttajan hyötyfunktion ja kuluttajan sijoittumiseen liittyvät muuttujat. Asuntojen hedonisten hintojen tapauksessa kaikista asuntojen hintoihin ja ominaisuuksiin liittyvistä tiedoista ei ole mahdollista saada laadukasta tilastoa. Lisäksi asuntojen ominaisuuksiin sisältyy useita multikollineaarisuuteen liittyviä ongelmia, kuten asuntojen ikä ja etäisyys keskustasta (vanhimmat asunnot on rakennettu keskustaan) ja asunnon kunto ja ikä. Jakaessa hintamallit alueellisesti voidaan välttää mallin estimoinnista usein esiin tulevat spesifikaatiovirheet ja toisaalta pienentää multikollineaarisuutta, koska riittävän tarkkoja malleja voidaan estimoida vähäisempien selittävien tietojen perusteella. Lopputuloksena segmentoitujen hintamalleja vertailemalla voidaan havaita eri alueilla vaikuttavat tekijät ja arvioida selittämättömän osan erot eri alueiden välillä. Tämän harjoitustyön tarkoituksena on tarkastella segmentoitujen hedonisten hintamallien avulla Helsingin asuntomarkkinoita: voidaanko hintamalleja käyttää aluetasolla ja mitkä tekijät selittävät asuntojen hintoja. Tarkastelussa hyödynnetään sekä havaintokohtaisia että postinumerotason tietoja (ks. Adair ym, 1996).

Normaalisti keskeiset kysymykset asuntomarkkinoiden eriytymistä koskevassa tutkimuksessa ovat olleet, onko eri alueiden asuntojen hintojen ja hintaerojen taustalla eri tekijöitä sekä voidaanko niitä selittää hedonisten hintamallien avulla. Samojen kysymysten pohjalta on tarkoitus tarkastella näitä kysymyksiä Helsingin aineistolla.

Harjoitustyön aineisto on poimittu Asuntojen.hintatiedot.fi –palvelusta, jota ylläpitää Asumisen rahoitus- ja kehittämiskeskus (ARA) sekä Ympäristöministeriö. Aineiston alueellisen ryhmittelyn pohjaksi päätettiin valita Tilastokeskuksen osakeasuntojen hintatilastossa käytettävä aluejako, jossa Helsinki on jaettu neljään osaan (ks. Liite 1, Tilastokeskus, 2015). Lisäksi aineistoa rikastettiin Tilastokeskuksen avoimilla postinumeroalueittaisilla tiedoilla (Tilastokeskus, 2016), joista hyödynnettiin postinumeroalueittaisia ansiotasotietoja.

Hintamallit on muodostettu usean muuttujan lineaarisella regressiomallilla käyttäen pienimmän neliösumman-menetelmää (PNS, eng. OLS). Lopullisen hintamallin valintakriteereinä on käytetty parasta selitysastetta sekä mallia, joka kaikista malleista minimoi ennustevirheiden neliösumman ja huomioi mallin muuttujien intuitiivisuuden.

## 2 Alueellisten hintamallien teoria ja aikaisempi tutkimus

Alueellisten asuntomarkkinoiden empiirinen tutkimus on hajautunutta, ja riippuen tarkastelun kysymyksenasettelusta - tutkimusten lähestymistavat eroavat toisistaan. Tässä esitetyt artikkelit ja niiden tulokset osoittavat asuntojen sijaintiin perustuvan ja asuntomarkkinoiden alamarkkinoiden tutkimuksen olevan hyvin moniselitteistä riippuen kaupunkialueen ja asuntomarkkinoiden ominaispiirteistä - muun muassa siitä mitä selittäviä tekijöitä on ollut käytössä.

Goodmanin ja Thibodeaun (1998) Yhdysvaltalaisella aineistolla tehdyssä tutkimuksessa hierarkkiset hintamallit mahdollistivat asuntomarkkinoiden alamarkkinoiden rajojen tarkastelun. Tutkimuksen perusteella asuntomarkkinoiden segregaatiota selitti julkisen koulutuksen laatuerot. Vastaavasti Goodmanin ja Thibodeaun (2003) myöhemmässä asuntomarkkinoiden eriytymistä koskevassa tutkimuksessa, jossa käytettiin sekä väestölaskennan tietoja että postinumeroaluejakoa, havaittiin paikkatietojen hyödyntämisen tuottavan merkittäviä hyötyjä hedonisille ennustemalleille.

Basun ja Thibodeaun (1998) dallasilaisella asuntomarkkina-aineistolla tehdyn tutkimuksen perusteella toteutuneiden hintojen ja ennusteiden erotus (residuaalit) olivat spatiaalisesti autokorreloituneita alamarkkinoiden mukaan kuudella alamarkkina-alueella kahdeksasta. Basu ja Thibodeau käyttivät hedonisten hintamallien estimoinnissa sekä pienimmän neliösumman menetelmää (PNS) että autoregressiivistä menetelmää (EGLS, estimated generalized least squares), jossa oletetaan regressiomallin residuaalien noudattavan tiettyä autoregressiivistä prosessia. Menetelmän avulla saavutetaan tarkempia ja tehokkaampia estimaatteja kuin PNS-estimoinnilla.

Archer et. al (1996) tarkastelivat asuntojen arvonnousua Miamin asuntomarkkinoilla. Tutkimuksen perusteella asuntojen hinnat ovat alueellisesti jokseenkin riippuvaisia mutta aluekohtaisilla eroilla (etäisyys keskustasta, perhekoot, etninen tausta) on vain vähäinen selitysvoima hintaerojen selittämisessä. Kontrolloimalla väestölaskentaryhmien kotien sijainnin huomattiin sen selittävän vain 12 % residuaalien vaihtelusta. Tutkimuksen perusteella asuinalue (*tract group*) ei juurikaan selitä hintaeroja tai nosta hintamallin selitystasetta.

Kielin ja Zabelin (2004) tutkimuksen perusteella alueellinen liikkumattomuus vaikuttaa asuntojen arvoon. Kolmen sijaintitekijän mallissa (*3Location -model*) kaikki kolme tekijää ovat tilastollisesti merkitseviä asuntojen hintaerojen kannalta: alueelliset klusterointitekijät, koulualueet ja alueellinen palvelujen tarjonta. Tutkimuksen tulokset ovat toisaalta ristiriitaisia, koska tutkimuksen perusteella asuntojen hintojen taustalla havaittiin olevan voimakas satunnaistekijä, joka aiheuttaa kolmeen tekijän estimaatteihin harhaa.

Edellisen tutkimuksen tulosten ja estimaattien harhaisuuden ja epästabilisuuden taustalla saattaa olla selittävien muuttujien korrelaatorakenteet ja se, että tiettyjen alueellisten tekijöiden laadukas mittaaminen on haasteellista.

Adairin ym. (1996) perusteella asuntojen alueelliset hintaerot johtuvat rakenteellisista ja sijaintikohtaisista tekijöistä sekä kysynnän ja tarjonnan lyhyen ajan epätasapainosta (vrt.

asuntomarkkinoiden kontrasyklisyys: tarjonta reagoi kysynnän muutoksiin viiveellä). Tutkimuksessa esitettiin, että osa asuntomarkkinoiden alueellisista eroista voi pysyä yli ajan mutta ei ole varmuutta mitä ne ovat. Adair ym. pyrkivät estimoimaan hintamalleja, joiden avulla voitaisiin havaita ja testata asuntomarkkinoiden segregatiota ja olemassaoloa. He havaitsivat, että makrotason hintamalleilla on hyvä ennustekyky, minkä takia asuntomarkkina-alueen ei tarvitse hintaerojen tarkastelun kannalta olla välttämättä pieni tarkasti määritelty homogeeninen alue. Lisäksi eri alamarkkinoiden malleissa oli eri selittävät tekijät, mikä heijastelee alueellisia eroja itsessään.

Seuraavassa luvussa alueellisia asuntomarkkinoita tarkastellaan Helsingin aineistolla mukaillen Adairin ym. (1996) tutkimusta.

## 3 Alueelliset hintamallit Helsingin aineistolla

### 3.1 Aineisto

Aineisto poimittiin Asuntojen.hintatiedot.fi –palvelusta, josta voi poimia toteutuneita asuntokauppätietoja viimeisen 12 kuukauden ajalta. Palvelua ylläpitää Asumisen rahoitus- ja kehittämiskeskus (ARA) sekä Ympäristöministeriö ja tietoja palveluun toimitetaan Kiinteistönvälitysalan Keskusliiton (KVKL) yhteistyösopimuksen perusteella.

#### 3.1.1 Aineiston muodostus ja rajaukset

Tutkimuksen peruskysymyksen ja rajauksen johdosta päädyttiin hyödyntämään Tilastokeskuksen asuntojen neljännesvuositilastossa käyttämää alueluokittelua (Tilastokeskus, 2015), jossa Helsinki on jaettu postinumeroiden perusteella neljään osaan (ks. liite 1). Muutkin aluejaot olisivat voineet olla mahdollisia, esimerkiksi jakamalla Helsinki keskustaan ja kehiin, neljän ilmansuunnan mukaan tai jakamalla Helsinki vain kahteen osaan. Kaikissa jaotteluissa joudutaan ratkomaan tiettyjä valintakriteereitä, jotka ovat enemmän tai vähemmän subjektiivisia. Tämän johdosta Tilastokeskuksen valmis luokittelu oli käyttökelpoisin. Näin ollen aineistoon poimittiin toteutuneet kerrostalojen asuntokaupat koko Helsingin alueelta niin, että muodostettiin suoraan neljä osa-aineistoa. Helsinki 1 pitää sisällään Helsingin kantakaupungin alueen ja niin sanotut arvoalueet. Helsinki 2 sisältää keskusta-alueen tuntumassa olevat lähialueet (mm. Lauttasaari, Takatöölö, Kallio ja Käpylä) mutta myös uusia asuinalueita ranta-alueilta (Herttoniemenranta, Aurinkolahti). Helsinki 3 –alue käsittää kaupunginosia hieman kauempaa keskustasta mutta arvostetuilta kerrostalo- ja omakotitaloalueilta ja usein hyvien liikenneyhteyksien varrelta. Helsinki 4 –alue on oikeastaan kaatoluokka, jossa on lukumääräisesti eniten Pohjois- ja Itä-Helsingin lähiöalueita.

Lisäksi asuntokauppakohtaiseen havaintoaineistoon lisättiin Tilastokeskuksen (2016) postinumeroaluetasoiset ansiotiedot. Vastaavaa tapaa on käytetty myös muissa tarkasteluissa (ks. esim. Diamond, 1980 ja Adair ym. 1996). Adairin ym. mukaan tämän tyyppiset taloudelliset ja sosioekonomiset muuttujat heijastelevat toisaalta alueen ja naapuruston laatua. Heidän Belfastia käsittelevässä tutkimuksessaan tämän tyyppisiä ns. proxyjä edustivat uskontokunnat ja niiden vaikutus hintoihin. Aineistoon olisi voinut lisätä myös muita postinumeroalueittaisia tietoja (esim. asukkaiden koulutusaste, työttömyys, työpaiikat toimialoittain) mutta tällöin multikollineaarisuus selittävien muuttujien välillä kasvaisi, ja toisaalta tiettyjen käyttökelpoisten tietojen läpikäynti ja jalostaminen malliin sopivaksi ei olisi mielekästä. Tällaisten postinumeroaluekohtaisten makromuuttujien laadukas tilastoaineisto ja tietojen riippuvuus keskenään, ja siitä seuraavat estimointiongelmien olivat nähtävissä myös asuntomarkkinoiden alueellisia hintaeroja selittävässä tutkimuksissa.

#### 3.1.2 Aineiston muuttujamuunnokset ja tunnusluvut

Aineiston poimimisen jälkeen aineistoon tehtiin mallien estimointien mahdollistamiseksi useita muuttujamuunnoksia. Muuttujamuunnokset on esitetty liitteessä 2. Aineiston havainnollistamiseksi osa luokitteluasteikollisista muuttujista mm. huoneistotyyppi, parveke ja sauna on jo muunnettu dummy –muuttujiksi (arvo 1 = kyllä, 0 = ei). Aineiston tarkas-

telun ja ensimmäisen estimointivaiheen jälkeen havaittiin, että asunnon rakennusvuodesta pääteltävä talon ikä ei ollut tilastollisesti merkitsevä missään malliversiossa. Näin ollen perustuen asunnon ikä-neliöhinta tarkasteluun ja yleiseen tietämykseen Helsingin rakentamisesta ikä –muuttuja jaettiin viiteen luokkaan.<sup>1</sup>

Asunnon toteutuneet hinnat muunnettiin alkuvaiheessa logaritmiselle asteikolle, koska tiedettiin myyntihinnan aiheuttavan heteroskedastisuutta ts. estimoitavan mallin residuaalien varianssin riippuu selittäjästä - esim. suuri selittäjän arvo ja samaan aikaan residuaalien suuri varianssi.<sup>2</sup> Velattoman myyntihinnan lisäksi huoneisto pinta-ala ( $m^2$ ) ja postinumeroalueen keskitulot muunnettiin logaritmiselle asteikolle. Logaritmoimalla velaton myyntihinta, pinta-ala ja keskitulot jatkuvat muuttujat toimivat tässä tapauksessa myös hyvin Box-Cox<sup>3</sup> menetelmän muunnoksina, joiden vaikutusta voidaan arvioida normaali-jakautuneisuutta testaavan tilastollisten testien perusteella sekä jakaumatunnuslukujen perusteella.<sup>4</sup> Shapiro-Wilk-testin perusteella kaikkien logaritimuunnosten jälkeen em. muuttujat ovat normaalijakautuneita. Lopullisesti muunnosten tehokkuutta voidaan arvioida parhaiten graafisesti siinä vaiheessa kun malli on estimoitu ja tarkastelemalla mm. residuaalien jakaumaa, hajontakuvion muotoa ja Q-Q-plotin lineaarisuutta (näiden tulokset on esitetty luvussa 3.2). Aineistoon lisättiin havainnoille yksilöivä surrogaatti id-tunnus mahdollisen havaintokohtaisen tarkastelun helpottamiseksi.

Mallien muodostamiseen käytetyt estimointiaineistojen jatkuvien muuttujien tunnusluvut on esitetty liitteessä 3. Tunnusluvuista on huomioitava logaritmoitujen muuttujamuunnosten lisäksi eri alueiden suuret erot neliöhinnoissa (keskineliöhinnat Helsinki 1: 6391 e/m<sup>2</sup> vrt. Helsinki 4: 2893 e/m<sup>2</sup>) mutta toisaalta neliömäärät ovat Helsinki 2 – 4 hyvin samalla tasolla (Helsinki 2: 54,5 m<sup>2</sup>, Helsinki 3: 57,6 m<sup>2</sup> ja Helsinki 4: 59,1 m<sup>2</sup>) mutta Helsinki 1:n keskimääräinen neliömäärä on 70,6 m<sup>2</sup>. Havainnoista suurin määrä oli Helsinki 2 –alueelta: 1 111 (42 % kaikista), Helsinki 4: 671 (25 %), Helsinki 1: 519 (20 %) ja Helsinki 3: 336 (13 %). Aluekohtaiset keskitulot eroavat myös selvästi Helsinki 1 ja muiden alueiden välillä (Helsinki 1: 36 260 euroa ja Helsinki 4 – Helsinki 2: 22 230 – 26 980 euroa).

---

<sup>1</sup> On huomioitavaa, että vuosikymmeniin jakaminen ei olisi optimaalista, koska rakentamisvuoden ja hinnan korrelaatio ei ole selkeää. Vanha asuntokanta jaettiin kahteen osaan: ennen vuotta 1920 valmistuneet ja vuodet 1921-1945, sodan jälkeinen jälleenrakennusvaihe ja ennen elementtitalorakentamista: vuodet 1946-1962, elementtitalovaihe ja lähiörakentaminen: vuodet 1963-1999 ja uusi asuntokanta vuoden 2000 jälkeen valmistuneet.

<sup>2</sup> Heteroskedastisuus näkyy myös residuaalien jakauman poikkeamana normaalijakautuneesta ja residuaalien odotusarvon poikkeamana nolasta. Mallin heteroskedastisuus johtaa PNS-estimaattorin tehottomuuteen (Gauss-Markov teoreema 4.), estimaattorit voivat olla harhattomia ja tarkentuvia mutta mallin heteroskedastisuus johtaa ennusteen epätarkkuuteen.

<sup>3</sup> Keskiarvoihin perustuvat tilastolliset analyysimenetelmät edellyttävät miltei aina normaalijakautuneita ja variansseiltaan samansuuruisia. Box-Cox –menetelmän avulla voidaan valita muunnos siten, että yhtä aikaa on saavutettavissa normaalijakautuneisuus, virhevarienssin vakioisuus ja mallin rakenteen yksinkertaisuus (lineaarisuus).

<sup>4</sup> Normaalijakautuneisuutta voidaan testata Shapiro-Wilk tai Kolmogorov-Smirnov –testien perusteella. Jakaumatunnusluvut skewness kuvaa jakauman vinoutta ja kurtosis jakauman huipukkuutta.



### 3.2 Mallin estimointi

Mallin spesifikaatioksi valitaan usean muuttujan lineaarinen regressiomalli<sup>5</sup>, jonka selittävä muuttuja on logaritminen velaton myyntihinta. Sovitettava malli on siis muotoa:

$$Y_i = \beta_0 + \beta_1 X_1 \dots \beta_n X_n + u_i,$$

jossa  $Y$  on selitettävä muuttuja (ennustettava velaton kauppahinta),  $X_i$  selittävä (ennustava) muuttuja ja  $\beta_i$  selittävän muuttujan kulmakerroin sekä  $u_i$  residuaali (virhetermi).

Tarkastelun ja aiemman kokemuksen perusteella estimointimetodiksi valitaan Stepwise. Estimointimenetelmää käytettäessä on kuitenkin myös huolehdittava, että selittävät muuttujat eivät ole keskenään korreloituneita, ja näin ollen tarkasteltiin muuttujien korrelaatiomatriiseja sekä VIF- ja Tolerance-arvoja. Aineisto oli siis jaettu neljään osaan aluejakoon perustuen. Jokaiselle alueelle estimoitiin useita malleja, joista valittiin tilastollisesti paras malli, jossa olisi korkein selitysaste (adjusted  $R^2$ ) ja pienin keskineliövirhe (RSME). Lisäksi mallin jokaisen muuttujan pitäisi olla tilastollisesti merkitsevä 95 % luottamustasolla (ks. t –arvot ja p –arvot), eikä muuttujien välillä saisi olla haitallisia korrelaatorakenteita ts. multikolinearisuutta (VIF-arvot lähellä 1 ja Tolerance arvo korkeampi kuin 0,2).

Ensimmäisessä vaiheessa jokainen alue ja siitä estimoidut mallit käytiin läpi tarkastellen havaitaanko estimointiin vaikuttavia outlier-havaintoja JASP –ohjelman diagnostiikan avulla.<sup>6</sup> Lisäksi ensimmäisessä vaiheessa havaittiin, että Helsinki 4 –alueen mallin selitysaste on heikko. Tämän johdosta päätettiin testiluonteisesti yhdistää alueiden 3 ja 4 aineistot, yhdistäminen ei kuitenkaan parantanut tuloksia joten tarkastelu jätettiin pois analyysivaiheesta.

Tämän jälkeen mallit estimoitiin uudelleen. Taulukkoon 1 on koottu alueittain parhaiden ja intuitiivisesti soveltuvimpien mallien tulokset vertailun helpottamiseksi. Taulukossa Helsinki 1 –alueen selitysaste (adjusted  $R^2$ ) on korkein (0,909), Helsinki 2 –alueen 0,861, Helsinki 3 –alueen 0,781 ja Helsinki 4 –alueen 0,673. Selitysasteiden perusteella voidaan esittää, että hedoninen malli soveltuu hyvin selittämään alueiden 1 ja 2 hintoja. Vastaa- vasti alueen 4 osalta malli soveltuu heikosti selittämään hintoja, joten käytössä olevien tietojen ja aluejako heijastuvat hinnoissa heikosti. Alueen 3 osalta tulosta voidaan pitää tyydyttävänä. Keskineliövirheet eivät eroa eri alueiden välillä merkittävästi.

Seuraavaksi varsinainen aluekohtainen hintamallivertailu perustuu mallien selittävien muuttujien kertoimien (vaikutuspainojen) vertailuun. Soveltuvimmiksi valituissa aluekohtaisissa hintamalleissa on eri määrä muuttujia. Helsinki 1 –alueen hedonisessa mallissa on kuusi muuttujaa, Helsinki 2:n mallissa kahdeksan, 3:n mallissa vain viisi, kun vastaavasti Helsinki 4:n mallissa kymmenen. Kaikki muuttujat ovat tilastollisesti selvästi

---

<sup>5</sup> Mallin estimoinnissa käytettiin JASP –tilasto-ohjelmaa, koska ko. ohjelmalla voidaan estimoida lineaarisen regressiomalli samoja estimointimetoodeja (ns. Enter, Forward, Backward, Stepwise -metodit) käyttäen kuin esim. SPSS –ohjelmalla.

<sup>6</sup> Jokaiselta alueelta poistettiin 1-2 suurinta outlier –havaintoa, jos residuaali toteutuneen ja ennustetun myyntihinnan perusteella erosi hyvin selvästi.

merkitseviä. Helsinki 4 -alueen heikon selitysasteen lisäksi mallin muuttujien korkea lukumäärä heijastelee osittain alueen heterogeenisuutta.

Muuttujakohtaisen tarkastelun perusteella havaittiin muutamia tekijöitä, jotka vastaavat ennakkokäsitystä:

- Kaikkiin malleihin valikoitui muuttujiksi huoneiston pinta-ala (m2\_log). Asunnon neliömäärän on tilastollisesti myös paras selittävä tekijä jokaisella alueella.
- Huoneiston hyvä kunto (KuntoHyva\_01) nousee malliin alueilla 1, 2 ja 4. Asunnon hyvä kunto vaikuttaa positiivisesti hintaan lähes yhtä voimakkaasti jokaisella alueella. Alueella 3 ja 4 myös asunnon huono kunto on nousee malliin selittäväksi tekijäksi.
- Keskitulo (Keskitulot\_log) nousee selittäväksi muuttujaksi kaikilla muilla alueilla paitsi Helsinki 3:lla.
- Rakennusvuoteen perustuvien dummy –muuttujia nousee jokaisen alueen malleihin ja ne ovat tärkeitä selittäviä tekijöitä.
  - Alueiden 1, 2 ja 4 –alueiden malleihin nousee elementtitalorakentamisen ja lähiörakentamisen aikajakso 1963-1999. Sen vaikutus hintaan on selvästi negatiivinen ja se on tilastollisesti merkitsevä. Tämä vastaa ennakkokäsitystä sikäli, ajanjakson rakentamista on pidetty heikkolaatuisena ja tuon ajan taloihin kohdistuu tällä hetkellä suurin osa peruskorjauksista.
  - Ajanjakson 1945-1963 vaikutus on alueen 2 mallissa negatiivinen ja alueen 3 mallissa positiivinen. Vaikutuksen ristiriitaisuutta voidaan pitää ennakoituna ja vaikutusta voidaan arvioida suhteessa alueen muuhun kantaan. Alueella 2 on paljon vielä vanhempia taloja, jotka sijaitsevat lähellä keskustaa. Vastaavasti alueella 3 sodan jälkeen rakennetut talot sijaitsevat arvostetuilla alueilla suhteessa muuhun alueen 3 rakennuskantaan.
  - Uutta rakennuskantaa kuvastavan ajanjakson (vuodet 2000-2016) vaikutus on alueilla 3 ja 4 selvästi positiivinen ja ne ovat tilastollisesti merkitseviä.
  - Ennen vuotta 1945 rakennusvuoden perusteella tehtyjä dummy-muuttujia ei nouse mihinkään malliin. Tämä johtuu todennäköisesti siitä, että alueen 1 kohdalla iso osa rakennuksista on tuolta aikakaudelta ja se ei ole keskiarvoihin perustuvassa analyysimenetelmässä tällöin tilastollisesti merkitsevä. Jos malli rakennettaisiin koskemaan koko Helsingin asuntomarkkinoiden hintoja nousisi muuttuja todennäköisesti malliin.
- Alinkerros laskee hintaa alueen 1 mallissa, jota voidaan pitää ennakoituna tiiviisti rakennetulla keskusta-alueella.
- Saunan olemassaolo nousi malliin alueilla 1, 3 ja 4. Hissin olemassaolo nousee selvästi malliin vain alueella 2 ja parvekkeen dummy-muuttuja alueella 4. Tämänkaltaisten dummy-muuttujien nousu selittämään hintaa liittyy myös keskiarvoihin perustuviin analyysimenetelmiin; alueilla 4 parveke on keskiarvosta poik-

keava tekijä jonka vaikutus on tilastollisesti positiivinen (vastaavasti esim. hissi alueella 2).

Aluekohtaisten mallien residuaalien jakaumatunnusluvut, jakaumakuvat ja hajontakuviot on esitetty liitteessä 4. Niiden perusteella residuaalit ovat kaikissa tapauksissa normaali-jakautuneita (keskiarvot, jakaumakuvat ja Q-Q –hajontakuva). Alueen 2 osalta on nähtävissä, että residuaalit eivät ole täysin normaalijakautuneita ja voidaan todeta heikkoa autokorrelaatiota, tästä huolimatta estimaattoreita voidaan pitää valitussa mallissa harhattomina ja malli tilastollisesti merkitsevä (ks. taulukko 1). Lisäksi, vaikka alueen 4 selitysaste oli heikko, ei mallissa ole residuaalitarkastelun perusteella havaittavissa autokorrelaatiota.

Muuttuja	Helsinki 1	Helsinki 2	Helsinki 3	Helsinki 4
intercept	2.077 (10.392)	3.405 (30.702)	4.161 (98.056)	1.457 (4.177)
m2_log	0.837 (63.414)	0.799 (74.572)	0.626 (25.861)	0.418 (18.538)
KuntoHyva_01	0.062 (10.421)	0.040 (10.039)		0.045 (7.549)
KuntoHuono_01			-0.079 (-5.154)	-0.061 (-3.940)
Keskitulot_log	0.433 (10.030)	0.139 (5.402)		0.693 (8.720)
Alinkerros_01	-0.043 (-4.965)			
Sauna_01	0.071 (7.661)		0.050 (4.727)	0.062 (6.986)
Parveke_01				0.025 (4.006)
Hissi_01		0.018 (4.037)		
Kerros		0.004 (3.742)		
Kaksio_01		-0.023 (-5.796)		-0.024 (-4.173)
NelioS_01				0.036 (3.470)
Rv_1920	-	-	-	-
Rv_1921_1945	-	-	-	-
Rv_1946_1962		-0.035 (-7.629)	0.082 (9.623)	
Rv_1963_1999	-0.068 (-6.078)	-0.050 (-8.455)		-0.055 (-7.691)
Rv_2000_2016			0.121 (13.133)	0.066 (5.091)
R <sup>2</sup>	0.910	0.862	0.785	0.678
Adjusted R <sup>2</sup>	0.909	0.861	0.781	0.673
RMSE	0.065	0.064	0.064	0.070
df	518	1110	335	670
F	868.0	857.2	207.1	138.7
p	< .001	< .001	< .001	< .001

**Taulukko 1.** Mallien vertailu. Muuttujan kertoimen perässä suluissa t-arvo.

## 4 Johtopäätökset

Mallikohtaisen vertailun perusteella keskeisinä havaintoina voidaan pitää Helsinki 1 ja 2 –alueiden korkeaa selitysasetta sekä toisaalta alueen 4 heikkoa selitysasetta. Tämä heijastelee samaa kuin Adairin ym. (1996) havainto, jonka mukaan hyvin määritellyt riittävän pienet alueet toimivat parhaiten hintamalleja estimoitaessa ja regressiomallien selitysasete nousee, kun asuntomarkkina-alueen homogeenisuus kasvaa. Vastaavan tuloksen ovat osoittaneet aiemmin myös esim. DeLisle (1984), Eckert (1990) ja Schuler (1990). Residuaalien keskineliövirheiden (RMSE) perusteella alueellisten hedonisten hintamallien erot eivät näyttäytyä niin selvästi. Voidaan esittää, että kantakaupungin homogeenisemmän aluejaon lisäksi käytettävissä olevat muuttujat kuvaavat paremmin huoneistojen ominaisuuksia ja niiden heijastumista toteutuneisiin hintoihin ja toisaalta, jos taustalla on tuntemattomia tekijöitä, ne tulevat huomioiduksi epäsuoraan asunnon ominaisuuksissa tai postinumeron alueen tulotasossa. On huomioitava, että Helsinki 2 –alue, joka pitää sisällään yli 1 100 havaintoa, vaikuttaa postinumerotasolla heterogeeniseltä mutta usean selittävän muuttujan perusteella hintamallin selitysasete nousee verrattain korkeaksi.

Helsingin tulosten perusteella voidaan asia esittää myös käänteisesti: alueen 4 heikko selitysasete (ja osittain alueen 3) heijastelee sitä, että käytettävissä olevat tiedot eivät heijastele asuntojen toteutuneita hintoja. Alue on toisaalta heterogeeninen niin kokonaisuudessaan kuin mikrolokaatioidenkin osalta. Alueiden sisällä voi olla esimerkiksi tonttien omistussuhteiden, yhtiön iän ja kunnan välillä suuria eroja ja lisäksi tarkat mikrolokaatiot ja tilastoista piilossa olevat tekijät aiheuttavat asuntojen välille suuriakin eroja saman postinumeron alueen sisällä. Tällainen tekijä voi olla muun muassa metro- tai juna-aseman läheisyys, jolla voi olla pääasiallisesti positiivinen vaikutus hintaan mutta toisaalta sijainti hyvin lähellä asemaa voi vaikuttaakin hintaan negatiivisesti. Tämän perusteella voidaan osoittaa, että aluekohtaisten hedonisten hintamallien avulla voidaan tarkastella aluekohtaisia asuntomarkkinoiden eroja ja taustalla olevia tekijöitä ja selitysaseteet kasvavat niillä alueilla, jotka ovat tarkasti rajattu ja mahdollisimman homogeenisiä.

Alueen 4 hintamallin (ja osittain alueen 3) selitysaseteen heikkouden syitä olisi mahdollista tutkia tarkemmin tekemällä useampia (tai vaihtoehtoisia) aluejakoja eri osamarkkinoihin, minkä myötä mahdollisesti paljastuisi, onko kyseessä puutteet selittävässä tekijöissä vai ennemmin osamarkkinoiden heterogeenisuus. Mikäli mallien selitysasete kasvaisi jakamalla alueet 3 ja 4 vielä useampiin osamarkkinoihin, voitaisiin tätä pitää indikaationa alueiden homogeenisuuden lisääntymisestä uudessa aluejaossa. Päinvastainen tulos vahvistaisi epäilyjä selittävien muuttujien puutteellisuudesta. Useamman alueen osamarkkinajaolla olisi myös mahdollista tarkentaa Helsingin eri osamarkkinoiden maantieteellistä rajautumista.

## 5 Liitteet

### Liite 1: Helsingin aluejako

<b>Helsinki 1</b>	
10	Helsingin keskusta
12	Punavuori
13	Kaartinkaupunki
14	Kaivopuisto
15	Eira
16	Katajanokka
17	Kruununhaka
18	Ruoholahti
22	Jätkäsaari
26	Keski-Töölö
<b>Helsinki 2</b>	
20	Lauttasaari
21	Vattuniemi
25	Taka-Töölö
27	Pohjois-Meilahti
28	Ruskeasuo
29	Meilahden sairaala-alue
30	Pikku-Huopalahti
31	Kivihaka
32	Etelä-Haaga
33	Munkkiniemi
34	Kuusisaari-Lehtisaari
50	Sörnäinen
51	Etu-Vallila
52	Itä-Pasila
53	Kallio
54	Kalasadama
55	Vallila
56	Toukola Vanha Kaupunki
57	Kulosaari
58	Sörnäinen
59	Kaitalahti
61	Käpylä
81	Herttoniemi
85	Jollas
99	Aurinkolahti
<b>Helsinki 3</b>	
40	Pohjois-Haaga
43	Maununneva
44	Lassila

62	Metsälä Etelä-Oulunkylä
65	Veräjämäki
66	Länsi-Pakila
67	Paloheinä
68	Itä-Pakila
69	Torpparinmäki
73	Tapanila
78	Tapaninvainio
79	Viikki
80	Herttoniemi
83	Tammisalo
84	Laajasalo
95	Vartioharju
<b>Helsinki 4</b>	
Muut postinumerot	

## Liite 2: Aineiston muuttujat ja muuttujamuunnokset

Muuttuja	Muuttujan kuvaus/muuttujamuunnos
m2	Huoneiston pinta-ala
m2_log	Log(m2), pinta-alan logaritmimuunnos
Huone_lkm	Huoneiden lukumäärä
Yksiö_01	Jos yksiö niin arvo 1, muuten 0
Kaksio_01	Jos kaksio niin arvo 1, muuten 0
Kolmio_01	Jos kolmio niin arvo 1, muuten 0
NelioS_01	Jos neliö tai suurempi niin arvo 1, muuten 0
Parveke_01	Jos parveke niin arvo 1, muuten 0
Sauna_01	Jos sauna niin arvo 1, muuten 0
Parveke_01	Jos parveke niin arvo 1, muuten 0
Sauna_01	Jos sauna niin arvo 1, muuten 0
Vh_log	Log(VelatonHinta), Velattoman myyntihinnan logaritmimuunnos
Eur_m2_log	Log(Eur/m <sup>2</sup> ), Velattoman neliöhinnan logaritmimuunnos
Rv	Rakennusvuosi
Ika	2016-Rv
Rv_1920	Jos rakennettu ennen vuotta 1921 niin 1, muuten 0
Rv_1921_1945	Jos rakennettu 1921-1945 niin 1, muuten 0

Rv_1946_1962	Jos rakennettu 1946-1962 niin 1, muuten 0
Rv_1963_1999	Jos rakennettu 1963-1999 niin 1, muuten 0
Rv_2000_2016	Jos rakennettu vuoden 1999 jälkeen niin 1, muuten 0
Kerros	Huoneiston kerros
Kerros_max	Talon korkein kerros
Ylinkerros_01	Jos kerros ylin niin 1, muuten 0
Alinkerros_01	Jos kerros alin niin 1, muuten 0
Hissi_01	Jos hissi niin 1, muuten 0
KuntoHyva_01	Jos kunto = 'hyvä' niin 1, muuten 0
KuntoTyyd_01	Jos kunto = 'tyyd' niin 1, muuten 0
KuntoHuono_01	Jos kunto = 'huono' niin 1, muuten 0
Keskitulot	Postinumeroalueen vuoden keskitulot
Keskitulot_log	Log(Keskitulot), keskitulojen logaritimuunnos
Mediaanitulot	Postinumeroalueen vuoden mediaanitulot
Mediaanitulot_log	Log(Mediaanitulot), mediaanitulojen logaritimuunnos

### Liite 3: Aineiston tunnusluvut

Helsinki 1	Huone_lkm	m2	m2_log	Vh	Vh_log	Eur/m2	lka	Keskitulot	Keskitulot_log
Valid	519	519	519	519	519	519	519	519	519
Missing	0	0	0	-	0	-	0	-	0
Mean	2,335	70,64	1,79	437 700	5,586	6 391	80,55	36 260	4,554
Std, Error of Mean	0,05195	1,688	0,01002	10 460	0,009461	60	1,447	266	0,002956
Median	2	62	1,792	375 000	5,574	6 225	89	33 550	4,526
Mode	2	40	1,602	245 000	5,389	6 667	89	32 960	4,518
Std, Deviation	1,184	38,44	0,2283	238 200	0,2155	1 373	32,96	6 048	0,06735
Variance	1,401	1478	0,05211	5,674e + 10	0,04645	1,886e +6	1086	3,658e +7	0,004535
Skewness	0,9965	1,343	-0,01516	2	0,2665	1	-1,069	1	0,8088
Kurtosis	0,9081	2,227	-0,3607	4	-0,4676	2	0,5749	1	1,696
Minimum	1	13	1,114	122 000	5,086	3 026	1	22 080	4,344
Maximum	7	241	2,382	1,770e +	6,248	12 810	166	49 970	4,699

**Taulukko 2.** Alueen 1 jatkuvien muuttujien tunnusluvut.

Helsinki 2	Huone_lkm	m2	m2_log	Vh	Vh_log	Eur/m2	Ika	Keskitulot	Keskitulot_log
Valid	1111	1111	1111	1 111	1111	1 111	1111	1 111	1111
Missing	0	0	0	-	0	-	0	-	0
Mean	2,122	54,54	1,697	270 300	5,395	5 082	57,66	26 980	4,423
Std, Error of Mean	0,02988	0,7195	0,005624	3 782	0,005145	28	0,8323	173	0,002378
Median	2	49	1,69	237 000	5,375	5 047	60	25 140	4,4
Mode	2	31	1,491	225 000	5,352	5 000	78	23 190	4,365
Std, Deviation	0,9961	23,98	0,1875	126 100	0,1715	937	27,74	5 758	0,07926
Variance	0,9922	575,1	0,03514	1,589e + 10	0,02941	878 800	769,7	3,315e + 7	0,006283
Skewness	0,9091	0,9582	0,01856	2	0,6141	0	-0,578	2	1,481
Kurtosis	0,9585	0,6864	-0,5756	8	0,3596	1	-0,476	10	2,985
Minimum	1	15,5	1,19	99 500	4,998	2 438	1	20 610	4,314
Maximum	6	160	2,204	1,210e +	6,083	9 534	109	62 350	4,795

**Taulukko 3.** Alueen 2 jatkuvien muuttujien tunnusluvut.

Helsinki 3	Huone_lkm	m2	m2_log	Vh	Vh_log	Eur/m2	Ika	Keskitulot	Keskitulot_log
Valid	336	336	336	336	336	336	336	336	336
Missing	0	0	0	-	0	-	0	-	0
Mean	2,339	57,6	1,733	218 000	5,316	3 924	37,21	24 240	4,381
Std, Error of Mean	0,0507	1,095	0,008521	4 169	0,007415	47	1,231	195	0,002696
Median	2	54	1,732	207 600	5,317	4 016	43	23 280	4,367
Mode	2	47,5	1,677	250 000	5,398	4 000	1	21 770	4,338
Std, Deviation	0,9293	20,07	0,1562	76 420	0,1359	856	22,56	3 583	0,04941
Variance	0,8636	403	0,02439	5,840e +	0,01848	733 100	509	1,284e + 7	0,002441
Skewness	0,4878	0,6229	-0,3046	2	0,4689	0	-0,475	6	3,895
Kurtosis	-0,01706	0,2913	-0,1695	6	0,5713	-	0	-1,272	23,13
Minimum	1	22	1,342	108 000	5,033	1 636	1	21 770	4,338
Maximum	5	119,5	2,077	660 000	5,82	6 410	68	51 950	4,716

**Taulukko 4.** Alueen 3 jatkuvien muuttujien tunnusluvut.

Helsinki 4	Huone_lkm	m2	m2_log	Vh	Vh_log	Eur/m2	Ika	Keskitulot	Keskitulot_log
Valid	671	671	671	671	671	671	671	671	671
Missing	0	0	0	-	0	-	0	-	0
Mean	2,408	59,13	1,748	162 400	5,193	2 893	42,78	22 230	4,346
Std, Error of Mean	0,03598	0,7256	0,005682	1 935	0,00471	32	0,5399	70	0,001321
Median	2	57,5	1,76	152 000	5,182	2 745	47	21 750	4,337
Mode	2	55	1,74	120 000	5,079	2 500	52	19 860	4,298
Std, Deviation	0,9321	18,8	0,1472	50 120	0,122	825	13,99	1 815	0,03421
Variance	0,8688	353,3	0,02166	2,512e +	0,01489	680 500	195,6	3,296e + 6	0,00117
Skewness	0,6788	0,3768	-0,5048	1	0,5306	1	-1,221	1	0,8833
Kurtosis	0,8113	0,2208	-0,08473	3	0,2339	1	1,135	1	0,5994
Minimum	1	22	1,342	79 000	4,898	1 043	0	18 940	4,277
Maximum	6	128,5	2,109	436 000	5,639	6 500	85	28 160	4,45

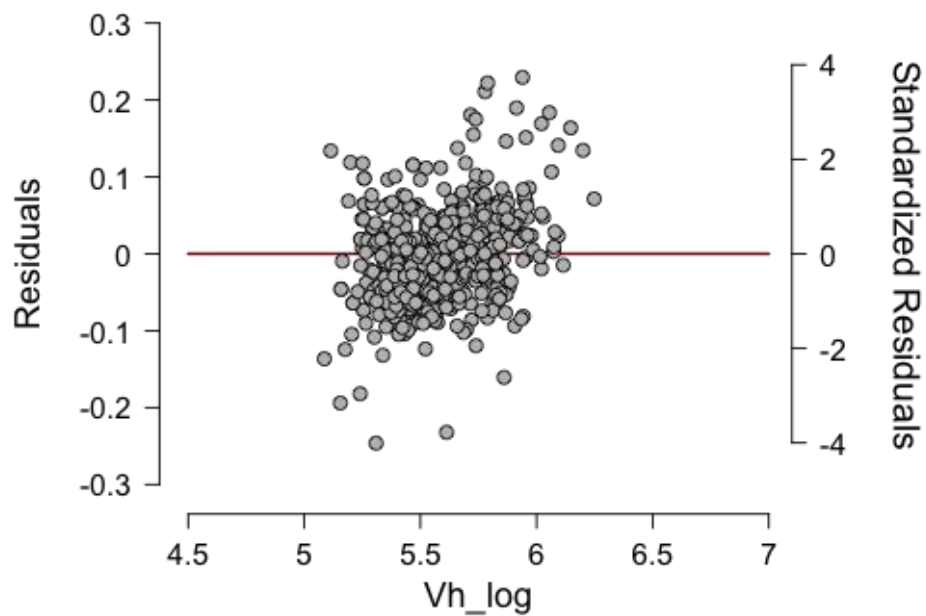
**Taulukko 5.** Alueen 4 jatkuvien muuttujien tunnusluvut.



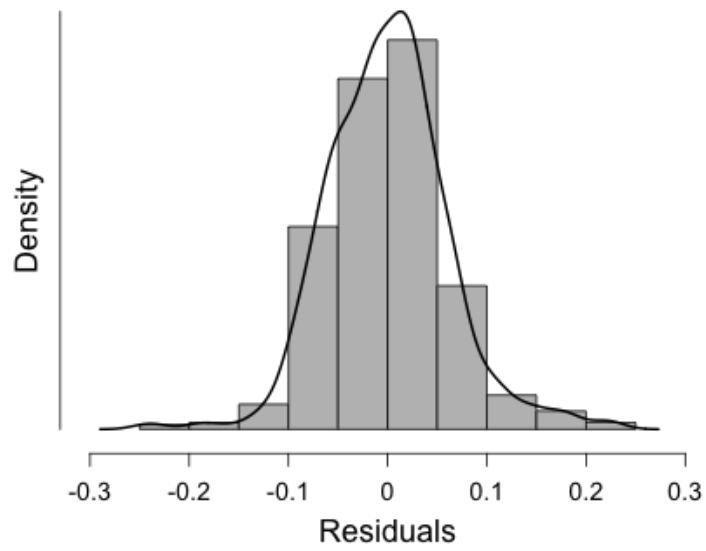
#### Liite 4: Mallien residuaalien diagnostiikka

Residuaalit	Helsinki 1	Helsinki 2	Helsinki 3	Helsinki 4
Miinimi	-0.246	-0.231	-0.176	-0.223
Maksimi	0.229	0.329	0.188	0.221
Keskiarvo	0.000	0.000	0.000	0.000
Keskihajonta	0.061	0.063	0.061	0.068
N	519	1111	336	671

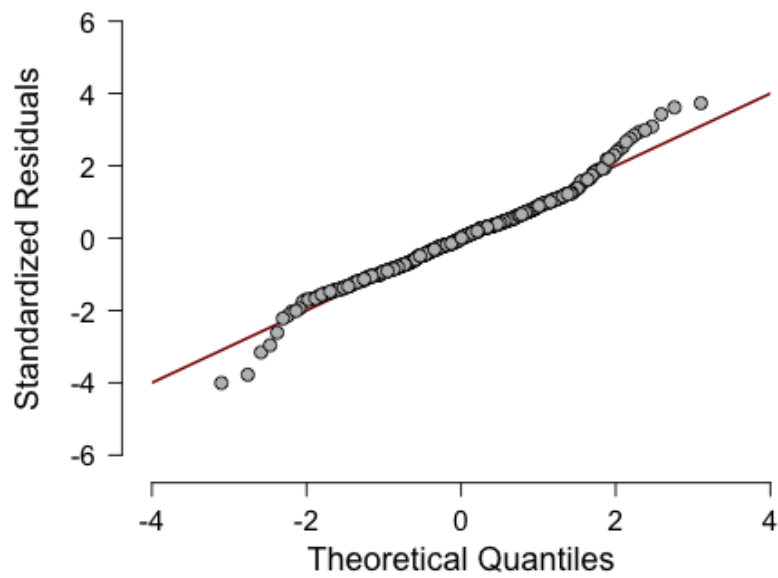
**Taulukko 6.** Mallien vertailu: Residuaalien tunnusluvut.



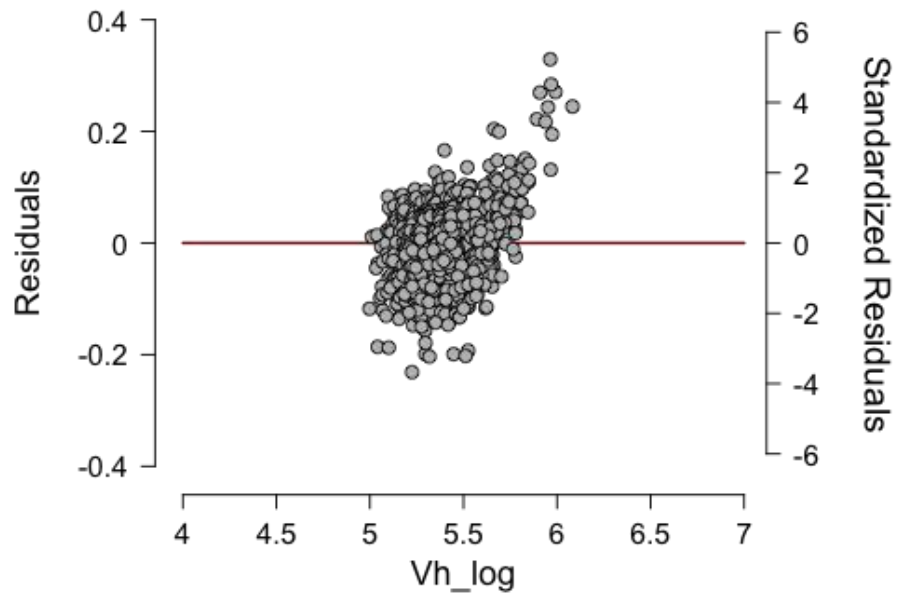
**Kuvio 1.** Alueen 1 residuaalien hajontakuvio suhteessa ennustettuun muuttujaan.



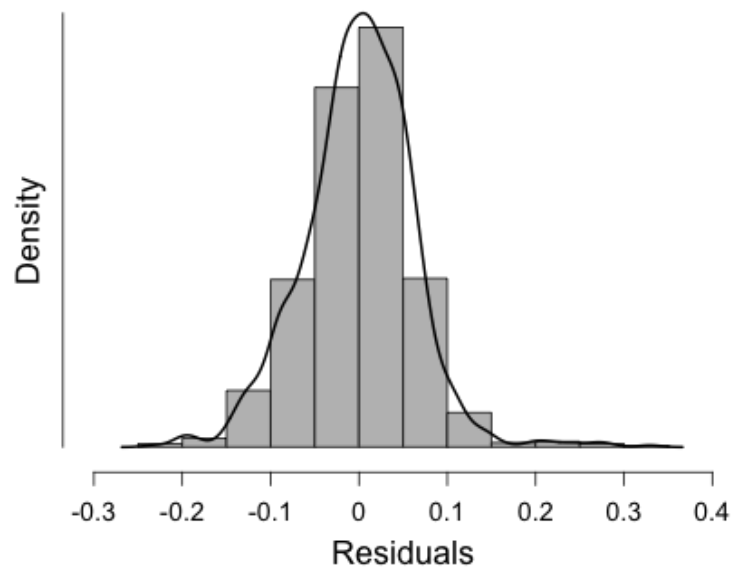
**Kuvio 2.** Alueen 1 residuaalien histogrammi.



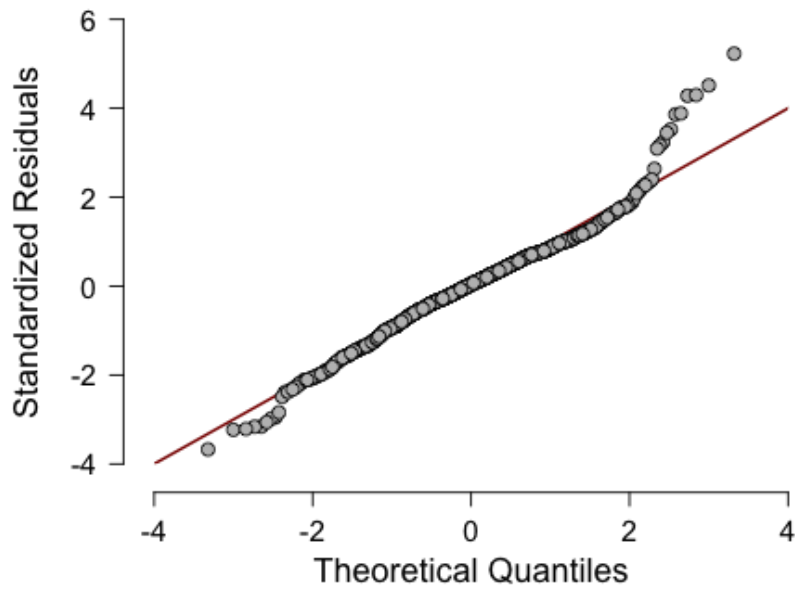
**Kuvio 3.** Alueen 1 standardoitujen residuaalien Q-Q -hajontakuvio.



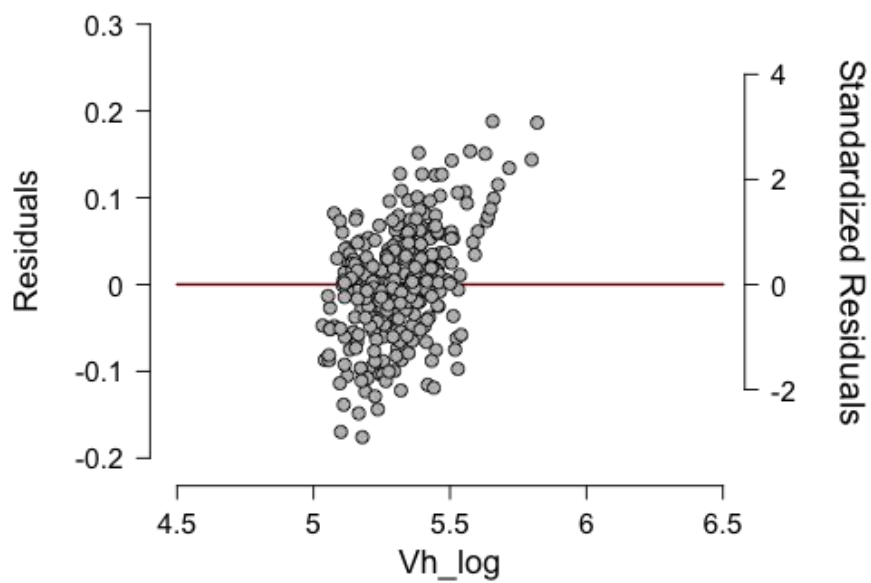
**Kuvio 4.** Alueen 2 residuaalien hajontakuvio suhteessa ennustettuun muuttujaan.



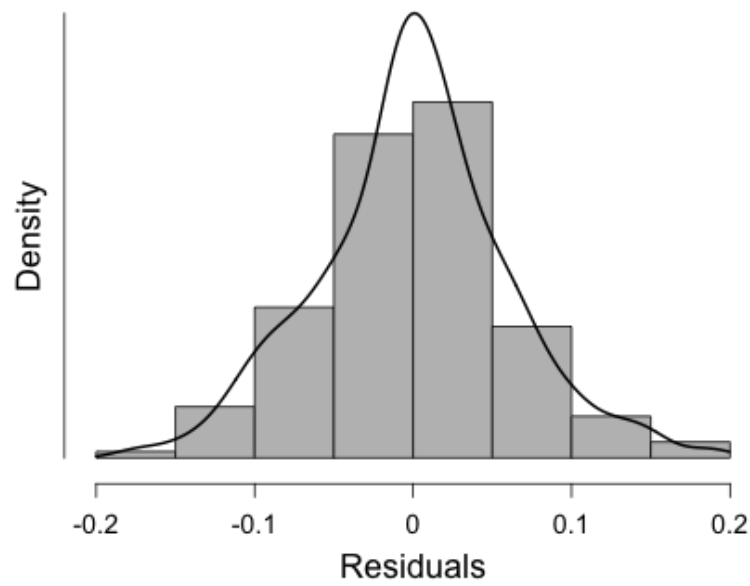
**Kuvio 5.** Alueen 2 residuaalien histogrammi.



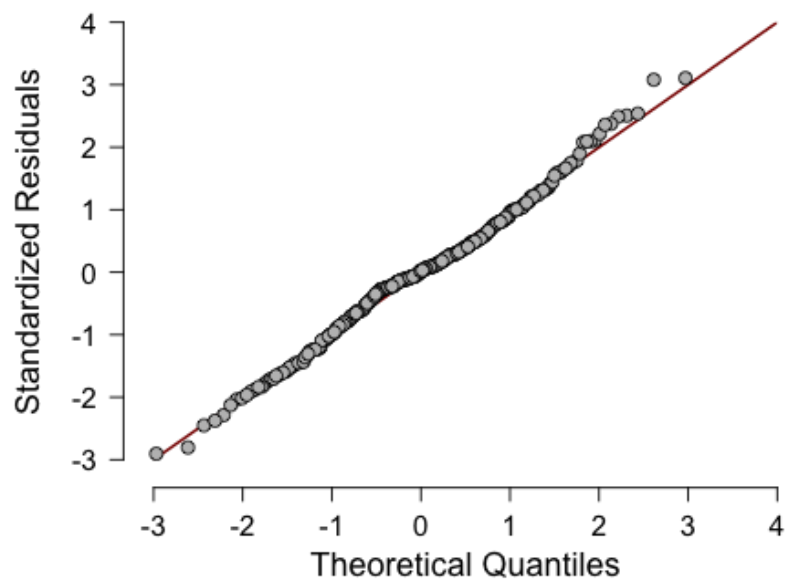
**Kuvio 6.** Alueen 2 standardoitujen residuaalien Q-Q -hajontakuvio.



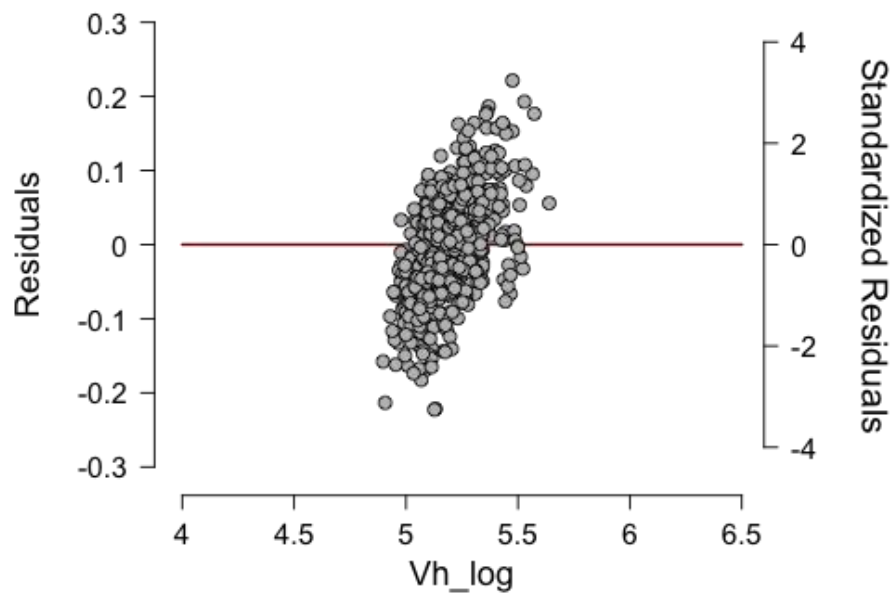
**Kuvio 7.** Alueen 3 residuaalien hajontakuvio suhteessa ennustettuun muuttujaan.



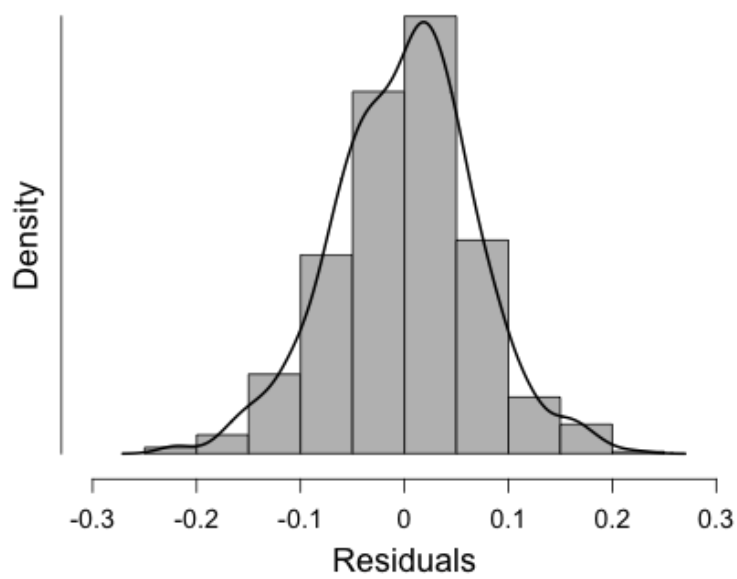
**Kuvio 8.** Alueen 3 residuaalien histogrammi.



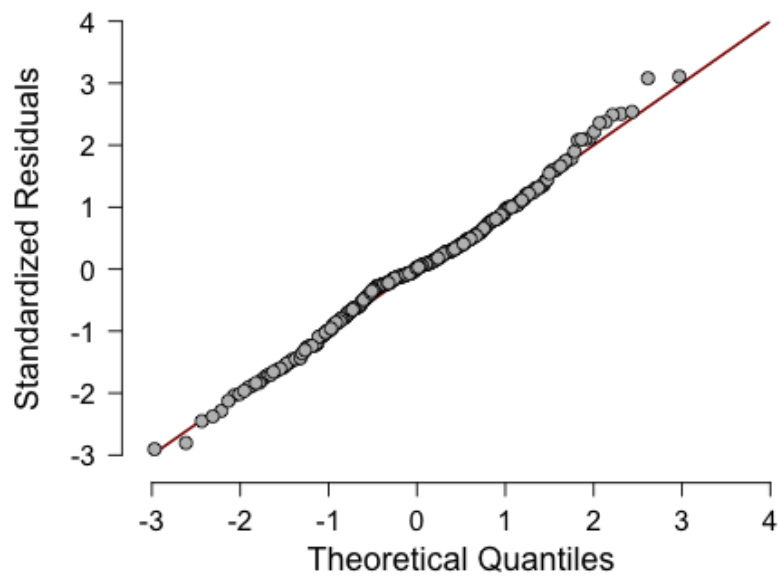
**Kuvio 9.** Alueen 3 standardoitujen residuaalien Q-Q -hajontakuvio.



**Kuvio 10.** Alueen 4 residuaalien hajontakuvio suhteessa ennustettuun muuttujaan.



**Kuvio 11.** Alueen 4 residuaalien histogrammi.



**Kuvio 12.** Alueen 4 standardoitujen residuaalien Q-Q -hajontakuvio.

## Lähdeluettelo

Adair, A. S., Berry, J. N. & McGreal, W. S. (1996). Hedonic modelling, housing submarkets and residential valuation. *Journal of Property Research*, 1996, 13, 67–83.

Archer, W. R., Gatzlaff, D. H., Ling, D. C. (1996). Measuring the importance of location in house price appreciation, *Journal of Urban Economics*, 40, 334-353.

Asuntojen.hintatiedot.fi (2016). <http://asuntojen.hintatiedot.fi/haku/>

Basu, S., Thibodeau, T., G. (1998). Analysis of spatial autocorrelation in house prices, *Journal of Real Estate Finance and Economics*, 17:1, 61-85.

DeLisle, J. R. (1984). Neighbourhood treatment in residential appraisal: a behavioural approach, *The Real Estate Appraiser and Analyst*, 50, 31–5.

Diamond, D. (1980). Income and residential location: Muth revisited. *Urban Studies*, 17, 1–12.

Eckert, J. K. (1990). *Property Appraisal and Assessment Administration*. International Association of Assessing Officers, Chicago.

Goodman, A. C., Thibodeau, T.G. (1998). Housing market segmentation. *Journal of Housing Economics* 7, 121–143.

Goodman, A. C., Thibodeau, T. G. (2003). Housing market segmentation and hedonic prediction accuracy, *Journal of Housing Economics*, 12, 181-201.

Kiel, K. A., Zabel, J. E. (2008). Location, location, location – the 3L approach to house price determination, *Journal of Housing Economics*, 17, 175-190.

Schuler, F.G. (1990). Econometric analysis of house prices, *The Real Estate Appraiser and Analyst*, 56, 20–32.

Tilastokeskus (2015). Osakeasuntojen hinnat. Aluejako neljännesvuositilastossa 29.4.2015. [http://www.stat.fi/til/ashi/ashi\\_2015-04-29\\_luo\\_001\\_fi.html](http://www.stat.fi/til/ashi/ashi_2015-04-29_luo_001_fi.html)

Tilastokeskus (2016). Postinumeroalueittainen avoin tieto. Asukkaiden käytettävissä olevat rahatulot 2013. <http://www.stat.fi/tup/paavo/index.html>